

GlyGen

Enhanced interface for retrieving glycan and
glycosylation data at GlyGen

Michael Tiemeyer, Sujeet Kulkarni, Robel Kashay, Rene Ranzinger, and
Raja Mazumder



PI: Michael Tiemeyer
CCRC, University of Georgia

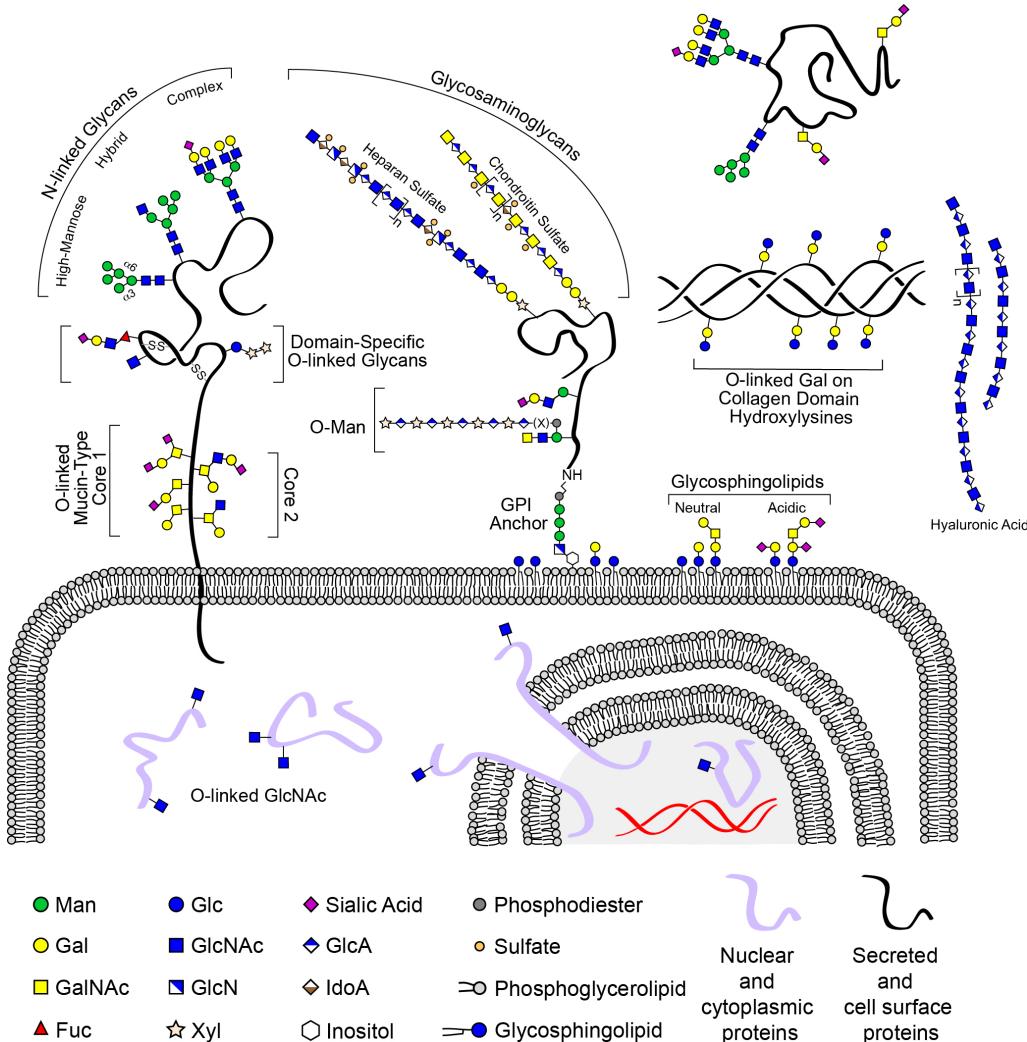


NIH Common Fund Glycoscience Program
1U01GM125267-01 (York & Mazumder)



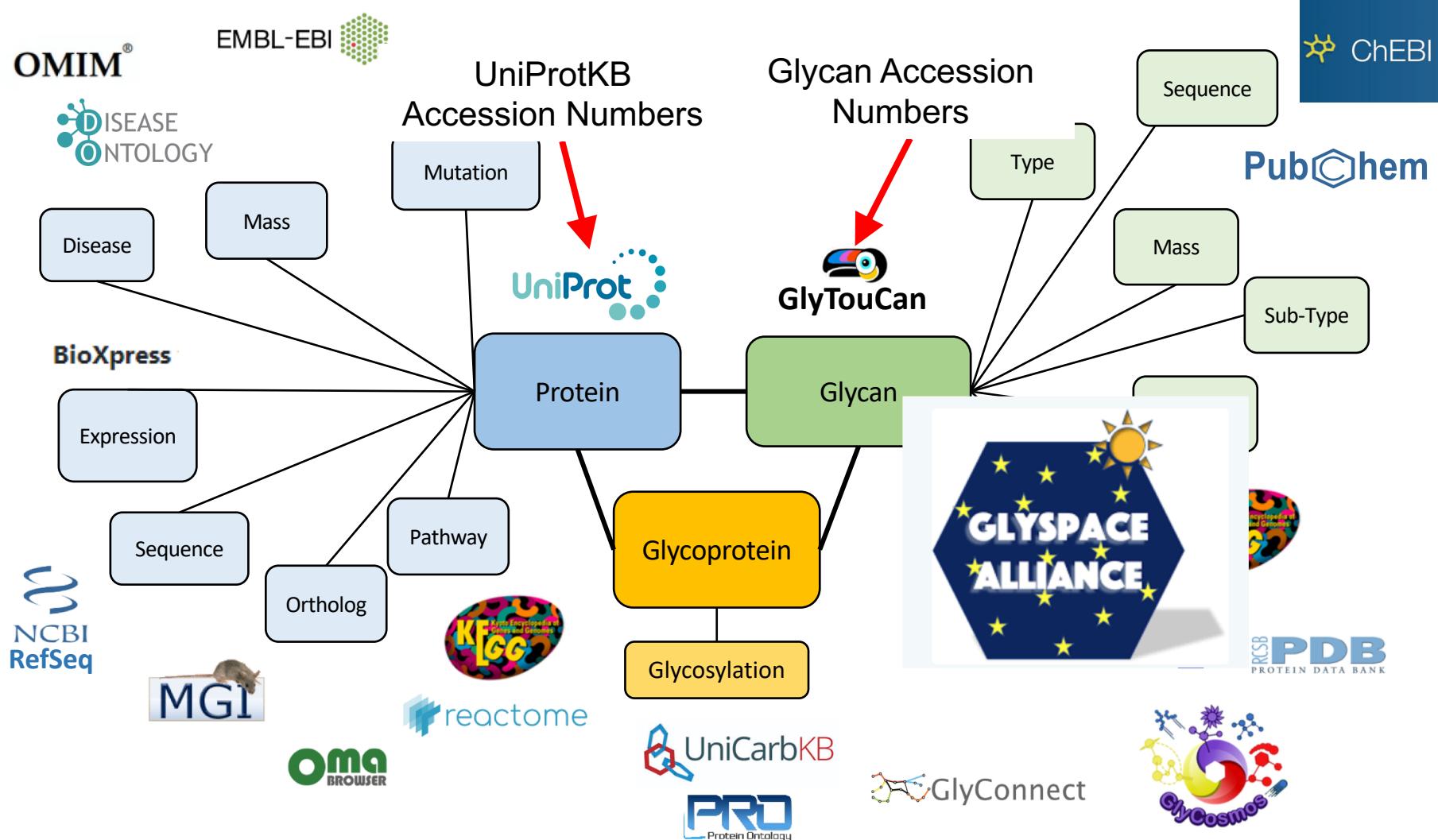
PI: Raja Mazumder
The George Washington University





- Glycoconjugates, whether secreted from cells or found at the surface or inside of cells, regulate essential functions in health and disease
- Almost all currently marketed biologic pharmaceuticals are glycoproteins (antibodies, growth factors, enzyme replacements)
- Glycan structural features influence glycoconjugate functions
- Where is this data captured? Can it be harvested in a way that is informative and that may reveal new knowledge?

Sources of data currently integrated into GlyGen



Introduction to GlyGen and spotlight a new feature of the most recent GlyGen release (v1.8, Apr. 2021)

- Introduction to Portal
 - Navigating and orienting, Try Me query
- GlyGen Super Search
 - Arriving at GlyGen with only a motif in mind
 - Acquiring novel answers to important questions

The Home Page

GlyGen

HOME EXPLORE ▾ QUICK SEARCH TRY ME DATA ▾ TOOLS ▾ HELP ▾ MORE ▾

Search...



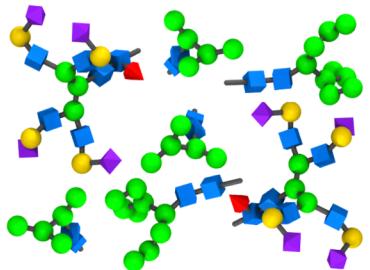
GlyGen: Computational and Informatics Resources for Glycoscience

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

HOW TO CITE

QUICK START

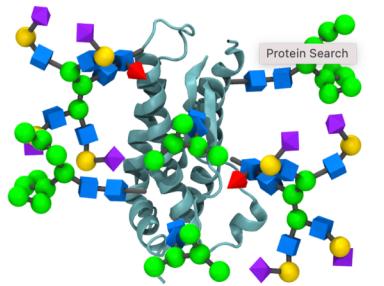
OUR MISSION



Glycan Search

Search for glycan structures based on their chemical and structural properties.

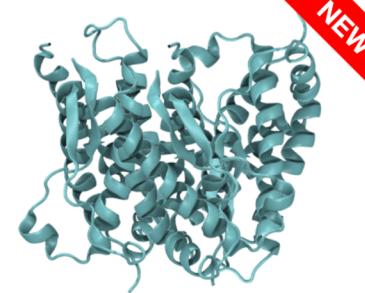
[EXPLORE](#)



Protein Search

Search for proteins based on their sequences, accessions, and annotations.

[EXPLORE](#)



Site Search

Search for proteins based on their site and site annotations.

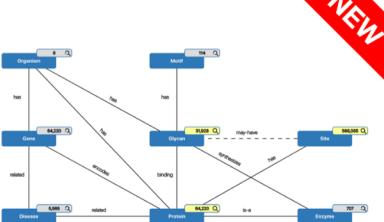
[EXPLORE](#)



Quick Search

Quick Search provides multi-domain queries that are based on user requests.

[EXPLORE](#)



Super Search

Super search is a graphical interface to build queries across all GlyGen datasets.

[EXPLORE](#)

GlyGen Mapper

ID mapping related to glycan, protein / glycoprotein and based on the user input.

[EXPLORE](#)

Version

Portal: 1.8 (19/Apr/2021)

Webservice: 1.8.48 (22/Apr/2021)

Data: 1.8.25 (19/Apr/2021)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

Database Statistics

Homo sapiens

Glycans	15855
Proteins	20609
Glycoproteins	9199

Mus musculus

Glycans	5842
Proteins	21989
Glycoproteins	3891

Rattus norvegicus

Glycans	4091
Proteins	21587
Glycoproteins	2111

Severe acute respiratory syndrome coronavirus 2

Glycans	1569
Proteins	17

[Scroll Down](#)

The Home Page

Composition Search
A search to find glycan(s) with specific monosaccharide composition in GlyGen.

[EXPLORE](#)

Structure Browser
Find glycan structures interactively by composition and topology via GNOme.

[EXPLORE](#)

List of Motifs
List of motifs includes detailed information and it's associated metadata.

[EXPLORE](#)

Glycoproteins	2
Hepatitis C virus (isolate H)	
Glycans	15
Proteins	1
Glycoproteins	1
Hepatitis C virus (isolate Japanese)	
Glycans	0
Proteins	1
Glycoproteins	1
Severe acute respiratory syndrome-related coronavirus	
Glycans	0
Proteins	15
Glycoproteins	3

Data
Data from the different resources can be accessed and downloaded in resource-specific formats (e.g. CSV, RDF).

[EXPLORE](#)

API
A public web service API allows access to the datasets by retrieving requested data in JSON format.

[EXPLORE](#)

SPARQL
All datasets are RDFized using standard ontologies and made accessible via a SPARQL Endpoint.

[EXPLORE](#)

Explore Other Resources

GlyGen is pleased to provide users with a variety of resources in glycobiology.

[EXPLORE](#)

News

GlyGen
@gly_gen

Attending [@ASMB](#) this week? Check out GlyGen's poster talk, Friday 4/30 at 3:30PM: "Enhanced interface for retrieving glycan and glycosylation data from GlyGen." M. Tiemeyer [@flyglyco](#) will present the latest strategies developed by the [@GlyGen](#) team for connecting data.

The Home Page

Explore this

GlyGen

HOME EXPLORE ▾ QUICK SEARCH TRY ME DATA ▾ TOOLS ▾ HELP ▾ MORE ▾

Search...



GlyGen: Computational and Informatics Resources for Glycoscience

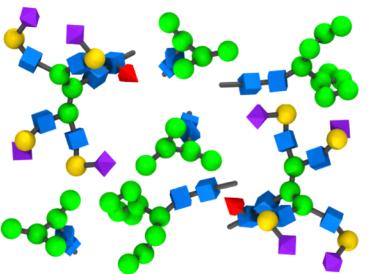
GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



HOW TO CITE

QUICK START

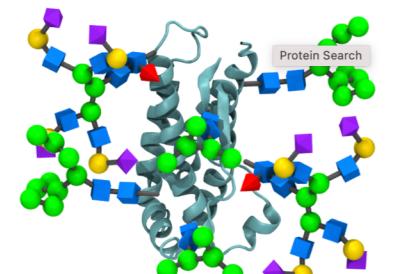
OUR MISSION



Glycan Search

Search for glycan structures based on their chemical and structural properties.

[EXPLORE](#)



Protein Search

Search for proteins based on their sequences, accessions, and annotations.

[EXPLORE](#)



Site Search

Search for proteins based on their site and site annotations.

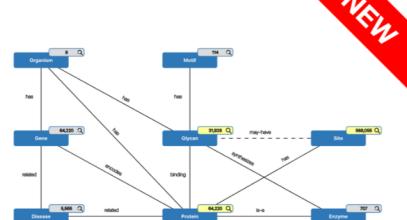
[EXPLORE](#)



Quick Search

Quick Search provides multi-domain queries that are based on user requests.

[EXPLORE](#)



Super Search

Super search is a graphical interface to build queries across all GlyGen datasets.

[EXPLORE](#)

GlyGen Mapper

ID mapping related to glycan, protein / glycoprotein and based on the user input.

[EXPLORE](#)

Version

Portal: 1.8 (19/Apr/2021)
Webservice: 1.8.48 (22/Apr/2021)
Data: 1.8.25 (19/Apr/2021)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

Database Statistics

Homo sapiens

Glycans	15855
Proteins	20609
Glycoproteins	9199

Mus musculus

Glycans	5842
Proteins	21989
Glycoproteins	3891

Rattus norvegicus

Glycans	4091
Proteins	21587
Glycoproteins	2111

Severe acute respiratory syndrome coronavirus 2

Glycans	1569
Proteins	17

Ready To Use Queries In Try Me

Try Me

What are the glycosyltransferases used to add the residues contained in [the lipid-linked N-glycan precursor GlcNAc2Man9Glc3 \(G72079JG\)](#) in human?

Which proteins have been shown to bear [a bi-antennary fully sialylated N-Glycan \(G77252PU\)](#) and which site is this glycan attached to?

Which glycans might have been synthesized in mouse using [Mgat1 \(P27808-1\)](#)?

Explore this



Protein Search Summary

Performed on: 2021/04/29 21:30:00

Which proteins have been shown to bear a bi-antennary fully sialylated N-Glycan (**G77252PU**) and which site is this glycan attached to?

[Update Results](#)[Modify Search](#)

** To perform the same search again using the current version of the database, click "[Update Results](#)".

Explore this

Records per page 20 ▾ Showing 1 to 16 of 16 Results

[DOWNLOAD](#) ▾

1

UniProtKB Accession ↑↓	Gene Name ↑↓	UniProt Name ↑↓	Hit Score ↑↓	Chemical Mass (Da) ↑↓	Organism ↑↓	RefSeq Name ↑↓	RefSeq Accession ↑↓
P07911-1	UMOD	Uromodulin, secreted form; Uromodulin	26.77	69761	Homo sapiens	uromodulin isoform a preproprotein	NP_001008390.1
P02787-1	TF	Serotransferrin	21.96	77064	Homo sapiens		
P05556-1	ITGB1	Integrin beta-1	21.41	88415	Homo sapiens	integrin beta-1 isoform 1A precursor	NP_002202.2
P0DN86-1	CGB3; CGB5; CGB8	Choriogonadotropin subunit beta 3	21.1	17739	Homo sapiens	choriogonadotropin subunit beta 3 precursor	NP_000728.1
P14210-1	HGF	Hepatocyte growth factor alpha chain; Hepatocyte growth factor beta chain; Hepatocyte growth factor	21.04	83134	Homo sapiens	hepatocyte growth factor isoform 1 preproprotein	NP_000592.3
P08648-1	ITGA5	Integrin alpha-5 heavy chain; Integrin alpha-5 light chain; Integrin alpha-5	20.6	114536	Homo sapiens	integrin alpha-5 precursor	NP_002196.4
P08519-1	LPA	Apolipoprotein(a)	20.17	501319	Homo sapiens		
P02788-1	LTF	Kaliocin-1; Lactoferricin-H; Lactoferroxin-A; Lactoferroxin-B; Lactoferroxin-C; Lactotransferrin	19.31	78182	Homo sapiens	lactotransferrin isoform 2	NP_001186078.1
P00533-1	EGFR	Epidermal growth factor receptor	18.96	134277	Homo sapiens	epidermal growth factor receptor isoform a precursor	NP_005219.2
P01215-1	CGA	Glycoprotein hormones alpha chain	18.52	13075	Homo sapiens	glycoprotein hormones alpha chain isoform 1 precursor	NP_001239312.1
P17050-1	NAGA	Alpha-N-acetylgalactosaminidase	18.38	46565	Homo sapiens	alpha-N-acetylgalactosaminidase precursor	NP_000253.1
P02743-1	APCS	Serum amyloid P-component(1-203); Serum amyloid P-component	16.57	25387	Homo sapiens	serum amyloid P-component precursor	NP_001630.1
Q03626-1	Mucl1	Murinolobulin-1	14.37	165326	Rattus	murinolobulin-1 precursor	NP_075591.1

Feedback



General
Glycosylation
Names
Function
Sequence
Single Nucleotide Variation
Mutagenesis
GO Annotation
Glycan Ligands
PTM Annotation
Proteoform Annotation
Pathway
Synthesized Glycans
Isoforms
Homologs
Disease
Expression Tissue
Expression Disease
Cross References
History
Publications

LOOK AT
Details For Glycoprotein P07911-1

Back

DOWNLOAD ▾



General

Gene Name: UMOD

Gene Location: Chromosome: 16 (20,356,301 - 20,333,052)

Ensembl Gene

UniProtKB ID: UROM_HUMAN

UniProtKB Accession: P07911-1

Protein Length: 640

UniProtKB Entry Name: Uromodulin, secreted formUromodulin

Chemical Mass: 69,761 Da

RefSeq Accession: NP_001008390.1

RefSeq Name: uromodulin isoform a preproprotein

Organism: Homo sapiens (Human) [9606]

UniProtKB



Glycosylation

ProtVista

Reported Sites with Glycan

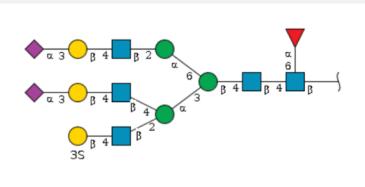
Reported Sites

Predicted Only

Text Mining

Records per page 20 ▾ Showing 1 to 20 of 718 Results

1 2 3 4 5 > >>

Source	Type ↑↓	GlyTouCan Accession ↑↓	Glycan Image	Residue ↑↓
UniCarbKB	N-linked	G03895SR		Asn 76
PubMed				

Types of data available on the detail page

Scroll Down

General
Glycosylation
Names
Function
Sequence
Single Nucleotide Variation
Mutagenesis
GO Annotation
Glycan Ligands
PTM Annotation
Proteoform Annotation
Pathway
Synthesized Glycans
Isoforms
Homologs
Disease
Expression Tissue
Expression Disease
Cross References
History
Publications

UniProtKB Entry Name: Uromodulin, secreted formUromodulin
Chemical Mass: 69,761 Da
RefSeq Accession: NP_001008390.1
RefSeq Name: uromodulin isoform a preproprotein
Organism: Homo sapiens (Human) [9606]

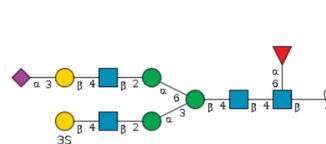
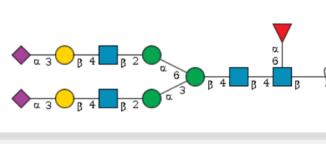
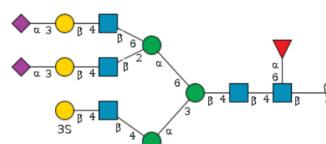
UniProtKB 1

Glycosylation

Reported Sites with Glycan Reported Sites Predicted Only Text Mining

Records per page 20 Showing 1 to 20 of 718 Results

1 2 3 4 5 > >>

UniCarbKB	PubMed	N-linked	Asn 76
1	2	G17540HT	
1	2	N-linked G17689DH	
1	2	N-linked G19469VZ	

Showing 1 to 20 of 718 Results

1 2 3 4 5 > >>

Explore this

Feedback

General
Organism
Names
Motifs
Associated Protein
Glycan Binding Protein
Biosynthetic Enzymes
Subsumption
Expression
Digital Sequence
Cross References
History
Publications

Types of data available on the detail page

LOOK AT

Details For Glycan **G17689DH**

Back

DOWNLOAD ▾

General

Sand Box Related Glycans ▲

GlyTouCan Accession: G17689DH
Monoisotopic Mass: 2,368.84 Da
Monoisotopic Mass-pMe (Da): 2,943.48 Da
Composition: Hex₅ HexNAc₄ dHex₁ NeuAc₂
Glycan Type / Glycan Subtype: N-linked / Complex
N-linked / Core-fucosylated
InCHI Key: YWJPMIROECFZJW-YBQNOACISA-N

Feedback

Introduction to GlyGen and spotlight a new feature of the most recent GlyGen release (v1.8, Apr. 2021)

- Introduction to Portal
 - Navigating and orienting, Try Me query
- GlyGen Super Search
 - Arriving at GlyGen with only a motif in mind
 - Acquiring novel answers to important questions

This paper may have captured your interest??

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 29, pp. 24356–24364, July 13, 2012
© 2012 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.

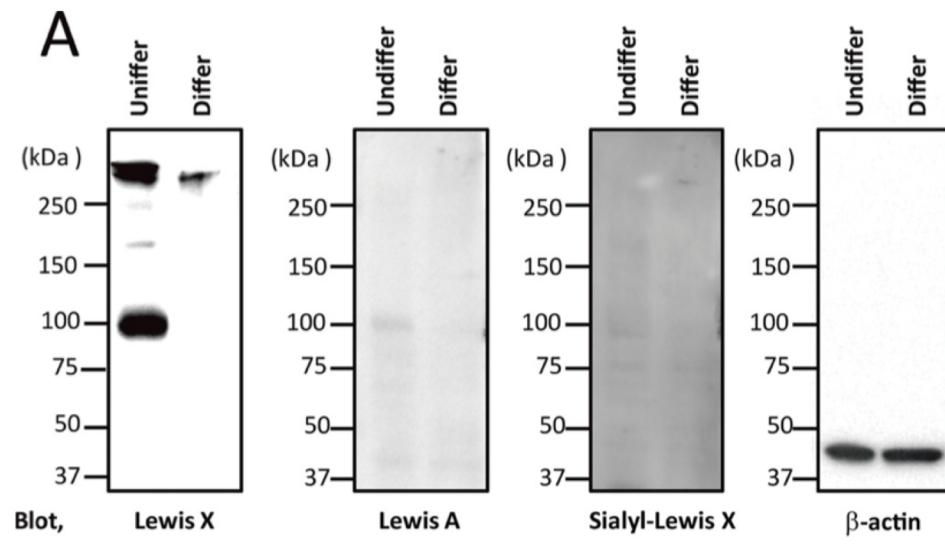
Lewis X-carrying N-Glycans Regulate the Proliferation of Mouse Embryonic Neural Stem Cells via the Notch Signaling Pathway^{*§}

Received for publication, March 23, 2012, and in revised form, May 10, 2012 Published, JBC Papers in Press, May 29, 2012, DOI 10.1074/jbc.M112.365643

Hirokazu Yagi^{‡1}, Takuya Saito[‡], Makoto Yanagisawa[§], Robert K. Yu^{§2}, and Koichi Kato^{‡¶3}

From the [‡]Graduate School of Pharmaceutical Sciences, Nagoya City University, 3-1 Tanabe-dori, Mizuho-ku, Nagoya 467-8603, Japan, the [§]Institute of Molecular Medicine and Genetics and Institute of Neuroscience, Medical College of Georgia, Georgia Health Sciences University, Augusta, Georgia 30912, and the [¶]Institute for Molecular Science and Okazaki Institute for Integrative Bioscience, National Institutes of Natural Sciences, 5-1 Higashiyama Myodaiji, Okazaki 444-8787, Japan

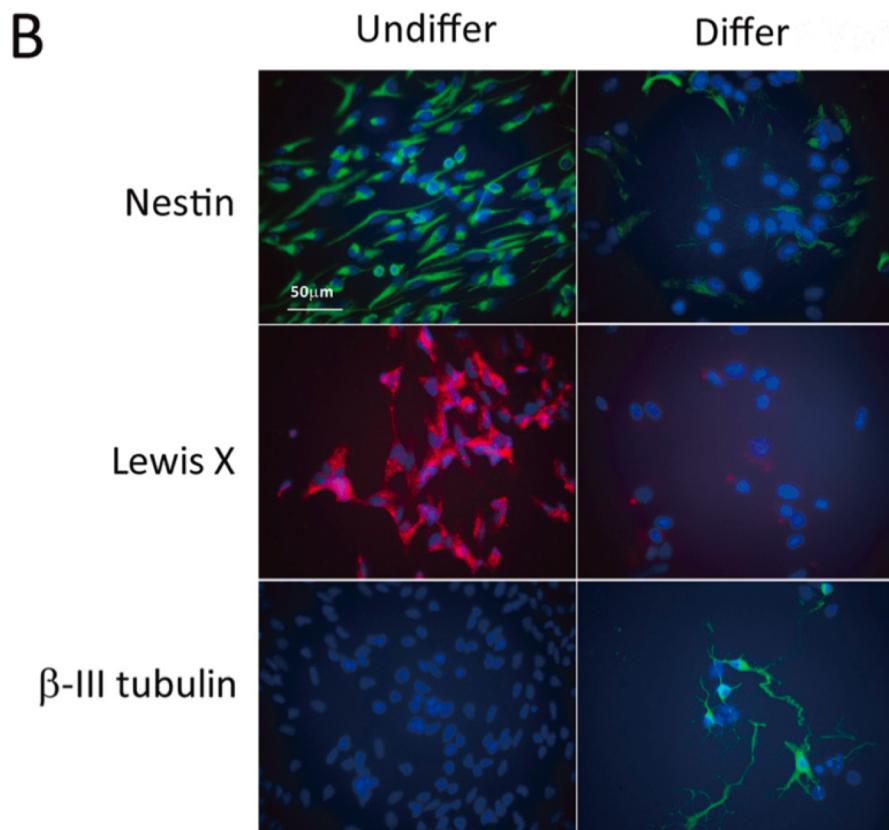
Undifferentiated stem cells express a much higher abundance of proteins that are modified with Lewis X glycans than differentiating neurons



What is Lewis X?

What proteins are modified with Lewis X?

Is Lewis X or are the proteins that carry it associated with any disease?



The Home Page

GlyGen

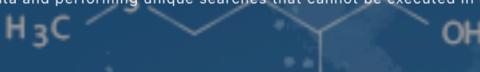
HOME EXPLORE ▾ QUICK SEARCH TRY ME DATA ▾ TOOLS ▾ HELP ▾ MORE ▾

Search...



GlyGen: Computational and Informatics Resources for Glycoscience

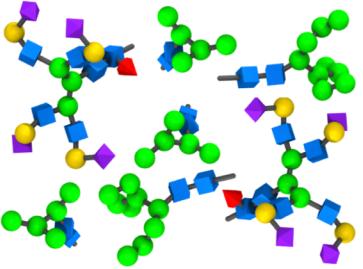
GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



HOW TO CITE

QUICK START

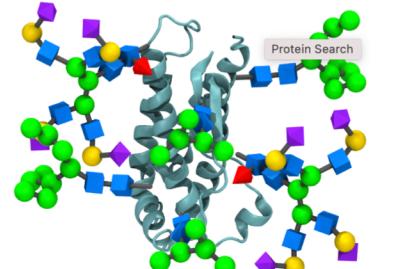
OUR MISSION



Glycan Search

Search for glycan structures based on their chemical and structural properties.

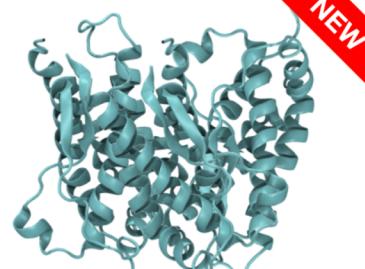
EXPLORE



Protein Search

Search for proteins based on their sequences, accessions, and annotations.

EXPLORE



NEW

Site Search

Search for proteins based on their site and site annotations.

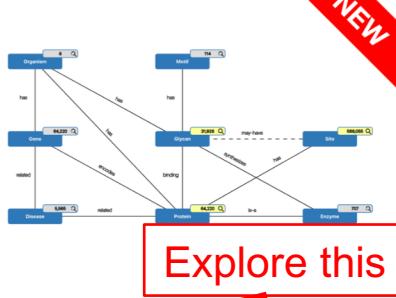
EXPLORE



Quick Search

Quick Search provides multi-domain queries that are based on user requests.

EXPLORE



Explore this

Super Search

Super search is a graphical interface to build queries across all GlyGen datasets.

EXPLORE

GlyGen Mapper

ID mapping related to glycan, protein / glycoprotein and based on the user input.

EXPLORE

Version

Portal: 1.8 (19/Apr/2021)
Webservice: 1.8.48 (22/Apr/2021)
Data: 1.8.25 (19/Apr/2021)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

LEAVE FEEDBACK

Database Statistics

Homo sapiens

Glycans	15855
Proteins	20609
Glycoproteins	9199

Mus musculus

Glycans	5842
Proteins	21989
Glycoproteins	3891

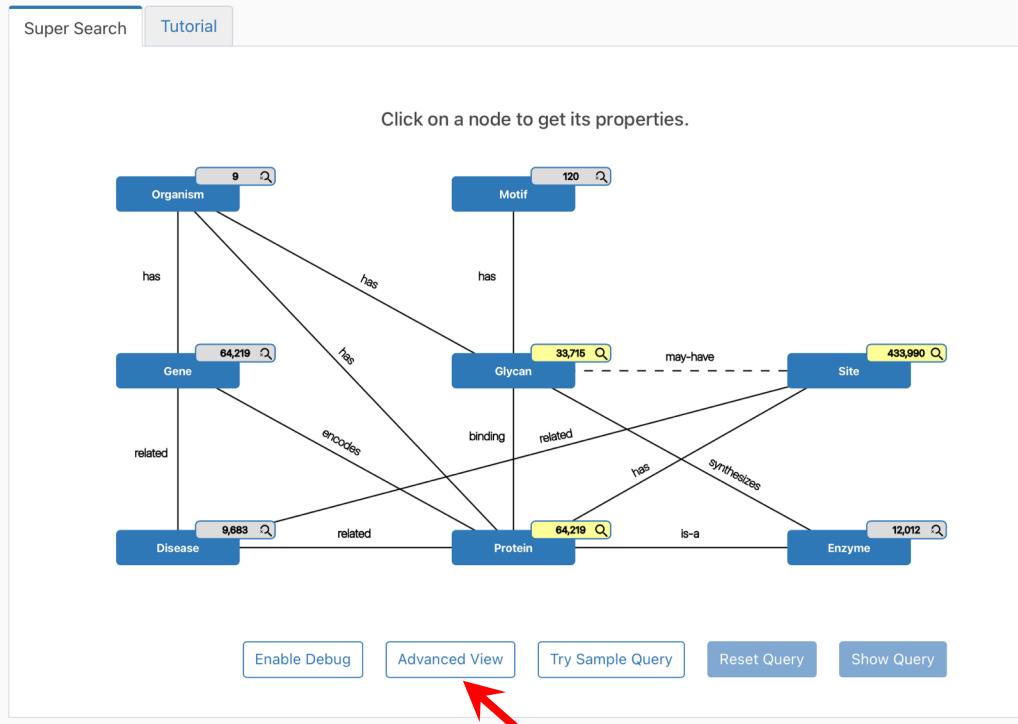
Rattus norvegicus

Glycans	4091
Proteins	21587
Glycoproteins	2111

Severe acute respiratory syndrome coronavirus 2

Glycans	1569
Proteins	17

Super Search

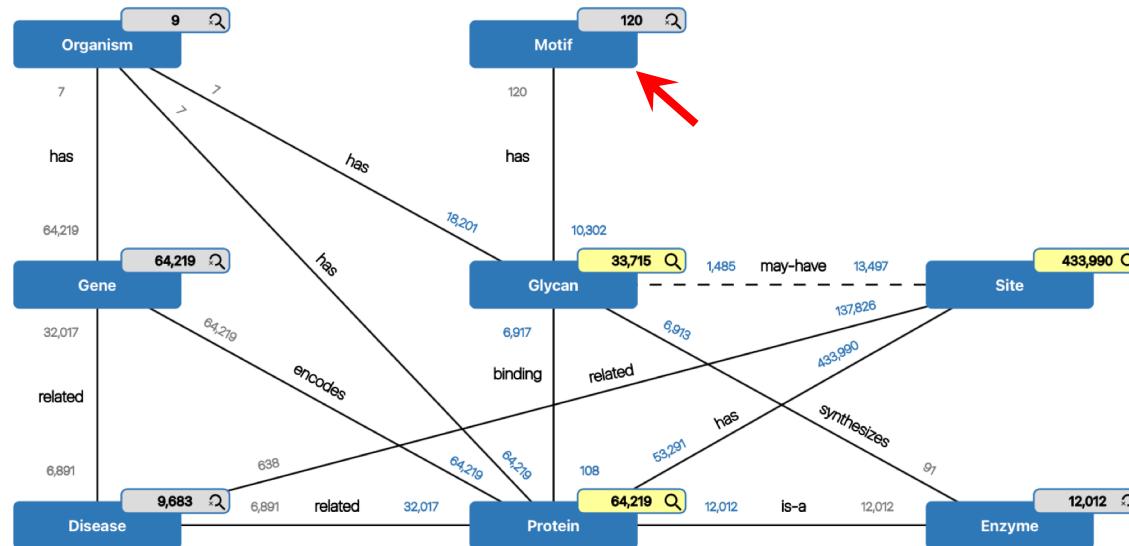


Super Search

Super Search

Tutorial

Click on a node to get its properties.



Enable Debug

Simple View

Try Sample Query

Reset Query

Show Query

Super Search

Super Search Tutorial

Click on a node to get its properties.

Organism Motif

Add motif properties to search

+ - ↴

- - ↴

- - ↴

- - ↴

Cancel Clear Fields Search

Super Search

Super Search

Tutorial

Click on a node to get its properties.

Organism



120



Motif

Add motif properties to search

GlyTouCan Accession



Motif ID



Publication Evidence (ID)



Publication Reference (Type)



Alignment_method



Publication Evidence (Database)



Reducing_end



Publication (Title)

Aglycon

Publication Reference (ID)

Publication (Author)

Publication (Year)

Publication (Journal)

Name

Cancel

Clear Fields

Search



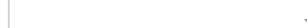
Super Search

Super Search [Tutorial](#)

Click on a node to get its properties.

Organism Motif

Add motif properties to search

Name	=	Lewis X	+  
			  
			
			  
			
<input type="button" value="Cancel"/> <input type="button" value="Clear Fields"/> <input type="button" value="Search"/>			



Super Search

Super Search Tutorial

Click on a node to get its properties.

The graph illustrates various biological relationships between entities:

- Organism** (5 nodes) has relationships with **Gene**, **Glycan**, and **Site**.
- Motif** (1 node) has a relationship with **Glycan**.
- Gene** (79 nodes) has relationships with **Glycan**, **Disease**, and **Protein**.
- Glycan** (1,649 nodes) is connected to **Organism**, **Motif**, **Site**, **Protein**, and **Enzyme**.
- Disease** (0 nodes) is connected to **Gene** and **Protein**.
- Protein** (86 nodes) is connected to **Organism**, **Motif**, **Glycan**, **Disease**, **Site**, and **Enzyme**.
- Site** (12 nodes) is connected to **Glycan** and **Enzyme**.
- Enzyme** (79 nodes) is connected to **Glycan**, **Protein**, and **Site**.

A red box labeled "Explore this" with an arrow points to the **Site** node.

Enable Debug Simple View Try Sample Query Reset Query Show Query

Site Search Summary

Performed on: 2021/04/29 21:45:12

Super search query.

[Update Results](#) [Modify Search](#)

** To perform the same search again using the current version of the database, click "Update Results".

Records per page: 20 Showing 1 to 12 of 12 Results 1

Explore this

UniProtKB Accession ↑↓	Hit Score ↑↓	Start Pos ↑↓	End Pos ↑↓	SNV	Glycosylation	Mutagenesis	Glycation	Phosphorylation
P01570-1	8.12	95	95	YES	YES	NO	NO	NO
P00450-1	8.12	138	138	YES	YES	NO	NO	NO
P02788-1	8.12	156	156	YES	YES	NO	NO	NO
P00450-1	8.12	397	397	YES	YES	NO	NO	NO
P00450-1	5.11	358	358	NO	YES	NO	NO	NO
P04925-1	5.11	196	196	NO	YES	NO	NO	NO
P02788-1	5.11	497	497	NO	YES	NO	NO	NO
P03956-1	5.11	120	120	NO	YES	NO	NO	NO
P19652-1	5.11	103	103	NO	YES	NO	NO	NO
P00450-1	5.11	762	762	NO	YES	NO	NO	NO
P04925-1	5.11	180	180	NO	YES	NO	NO	NO
O09164-1	5.11	121	121	NO	YES	NO	NO	NO

Showing 1 to 12 of 12 Results

1

General
Glycosylation
Names
Function
Sequence
Single Nucleotide Variation
Mutagenesis
GO Annotation
Glycan Ligands
PTM Annotation
Proteoform Annotation
Pathway
Synthesized Glycans
Isoforms
Homologs
Disease
Expression Tissue
Expression Disease
Cross References
History
Publications

LOOK AT

Details For Glycoprotein P01570-1

Back

DOWNLOAD ▾

Explore this

General

Gene Name: IFNA14

Gene Location: Chromosome: 9 (21,239,991 - 21,239,002)

Ensembl Gene 1

UniProtKB ID: IFN14_HUMAN

UniProtKB Accession: P01570-1

Protein Length: 189

UniProtKB Entry Name: Interferon alpha-14

Chemical Mass: 22,063 Da

RefSeq Accession: NP_002163.2

RefSeq Name: interferon alpha-14 precursor

Organism: Homo sapiens (Human) [9606]

UniProtKB 1

Glycosylation

ProtVista

Reported Sites with Glycan

Reported Sites

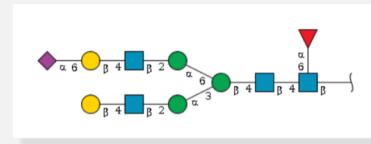
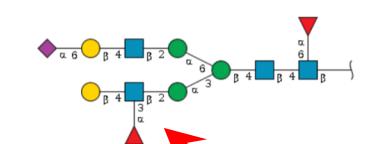
Predicted Only

Text Mining

Records per page 20

Showing 1 to 6 of 6 Results

1

Source	Type ↑↓	GlyTouCan Accession ↑↓	Glycan Image	Residue ↑↓
UniCarbKB 1 PubMed 1	N-linked	G07392QY		Asn 95
UniCarbKB 1 PubMed 1	N-linked	G75903TQ		Asn 95

Feedback

General
Glycosylation
Names
Function
Sequence
Single Nucleotide Variation
Mutagenesis
GO Annotation
Glycan Ligands
PTM Annotation
Proteoform Annotation
Pathway
Synthesized Glycans
Isoforms
Homologs
Disease
Expression Tissue
Expression Disease
Cross References
History
Publications

Observational study of gene-disease association. (HuGE Navigator)

RefSeq 1

Sequence

ProtVista

+10 +20 +30 +40 +50

1 MALPFALMMA LVVLSCK**S**SC SLGCNLSQL**H** SLNN**R**TLM**L** MAQMRRI**P**F SCLKD**R**HDFE
 61 FPQEEDGQNQ FQKAAQISV**L** HEMMQQTFLN FSTK**N**SSA P**E**Y**E**V**E**Y**E** FLCRQW**P**
 121 ACVIQEVG**V** ETPLMNEDSI LAV**K**YFQRI TL**L**YMEKK**Y** FCAWEVVRAE IMRSLSFSTN
 181 L**Q**KRLRRKD

N-linked Sites 1
 O-linked Sites
 Variation from Human 2
 Sequon 2

Explore this

Single Nucleotide Variation

ProtVista

Disease associated Mutations Non-disease associated Mutations

Records per page 20 Showing 1 to 7 of 7 Results

	rs1273950849.						
 BioMutant 1	Somatic mutation passed 1 out of 6 filters: n-glyco-sequon-loss (NSS->SSS). In dbSNP: rs1367674663.	Chr9:21239651	95	95	N → S	• colorectal cancer (DOID:9256)	
 dbSNP 1							
 gnomAD 1							
 BioMutant 1	Somatic mutation passed 1 out of 6 filters: patient freq. (1.4%).	Chr9:21239631	102	102	E → Q	• breast cancer (DOID:1612)	
 BioMutant 1	Somatic mutation passed 1 out of 6 filters: patient freq. (4.0%). In dbSNP: rs140823004.	Chr9:21239504	144	144	K → R	• prostate cancer (DOID:10283)	

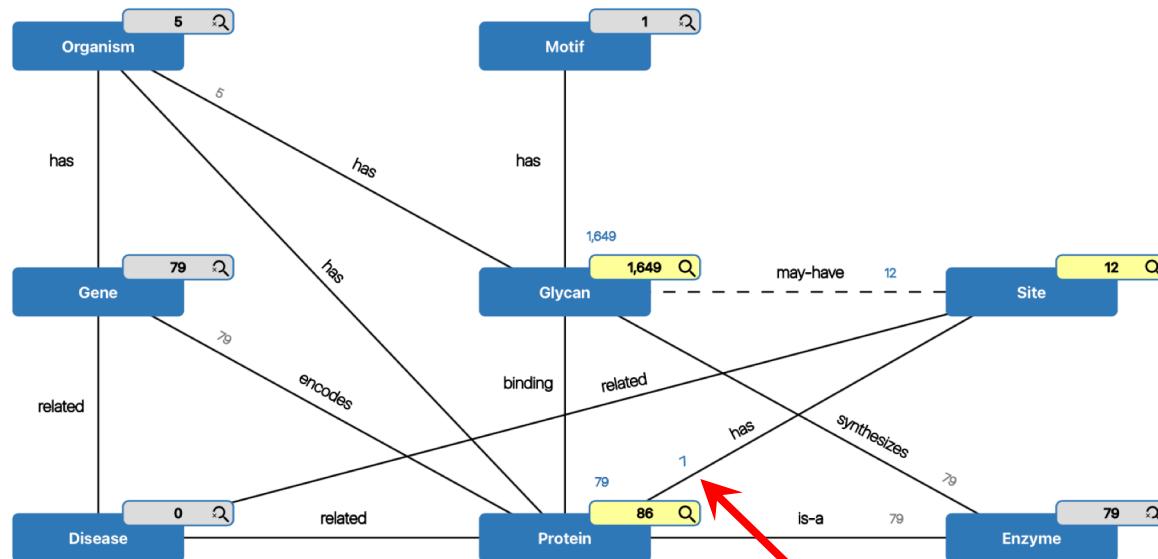
Showing 1 to 7 of 7 Results

Super Search

Super Search

Tutorial

Click on a node to get its properties.



Explore this

Enable Debug

Simple View

Try Sample Query

Reset Query

Show Query

Protein Search Summary

Performed on: 2021/04/29 21:57:42

Super search query.

[Update Results](#) [Modify Search](#)

** To perform the same search again using the current version of the database, click "["Update Results"](#)".

Records per page [20](#) Showing 1 to 7 of 7 Results [DOWNLOAD](#) 1 [Feedback](#)

UniProtKB Accession ↑↓	Gene Name ↑↓	UniProt Name ↑↓	Hit Score ↑↓	Chemical Mass (Da) ↑↓	Organism ↑↓	RefSeq Name ↑↓	RefSeq Accession ↑↓
P00450-1	CP	Ceruloplasmin	23.08	122205	Homo sapiens		
P02788-1	LTF	Kaliocin-1; Lactoferricin-H; Lactoferroxin-A; Lactoferroxin-B; Lactoferroxin-C; Lactotransferrin	19.31	78182	Homo sapiens	lactotransferrin isoform 2	NP_001186078.1
P19652-1	ORM2	Alpha-1-acid glycoprotein 2	13.22	23603	Homo sapiens	alpha-1-acid glycoprotein 2 precursor	NP_000599.1
P03956-1	MMP1	22 kDa interstitial collagenase; 27 kDa interstitial collagenase; Interstitial collagenase	12.02	54007	Homo sapiens	interstitial collagenase isoform 1 preproprotein	NP_002412.1
P04925-1	Prnp	Major prion protein	11.98	27977	Mus musculus	major prion protein precursor	NP_001265185.1
O09164-1	Sod3	Extracellular superoxide dismutase [Cu-Zn]	11.35	27392	Mus musculus	extracellular superoxide dismutase [Cu-Zn] precursor	NP_035565.1
P01570-1	IFNA14	Interferon alpha-14	11.29	22063	Homo sapiens	interferon alpha-14 precursor	NP_002163.2

Showing 1 to 7 of 7 Results 1

Introduction to GlyGen and spotlight a new feature of the most recent GlyGen release (v1.8, Apr. 2021)

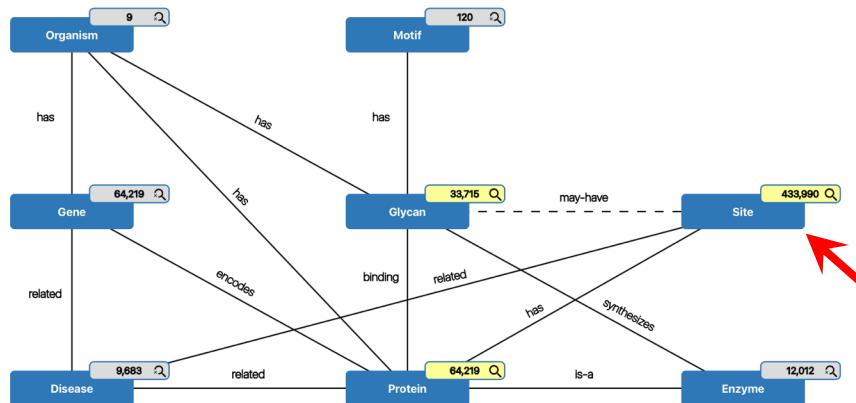
- Introduction to Portal
 - Navigating and orienting, Try Me query
- GlyGen Super Search
 - Arriving at GlyGen with only a motif in mind
 - Acquiring novel answers to important questions

Super Search

Super Search

Tutorial

Click on a node to get its properties.

[Enable Debug](#)[Advanced View](#)[Try Sample Query](#)[Reset Query](#)[Show Query](#)

Super Search

Super Search

Tutorial

Click on a node to get its properties.

Organism

9

120

Add site properties to search

Site Sequence



Start Position



End Position

Single Nucleotide Variation

Glycosylation

Mutagenesis



Cancel

Clear Fields

Search

Super Search

Super Search Tutorial

Click on a node to get its properties.

Organism Motif

Add site properties to search

Site Sequence = N + - ↓
And Glycosylation = true + - ↑
↓
And Single Nucleotide Variation = true + - ↑
↓

Cancel Clear Fields Search

The screenshot shows the 'Super Search' interface with a blue header bar. Below it, two nodes are shown: 'Organism' (9 results) and 'Motif' (120 results). A central message says 'Click on a node to get its properties.' The main area is titled 'Add site properties to search' and contains three rows of search fields. Row 1: 'Site Sequence' dropdown set to 'N'. Row 2: 'And' dropdown, 'Glycosylation' dropdown, and 'true' dropdown. Row 3: 'And' dropdown, 'Single Nucleotide Variation' dropdown, and 'true' dropdown. Each row has a '+' button, a trash bin button, and a downward arrow button. Red arrows point to the 'Glycosylation' field in the second row and the 'Single Nucleotide Variation' field in the third row. At the bottom are 'Cancel', 'Clear Fields', and 'Search' buttons.

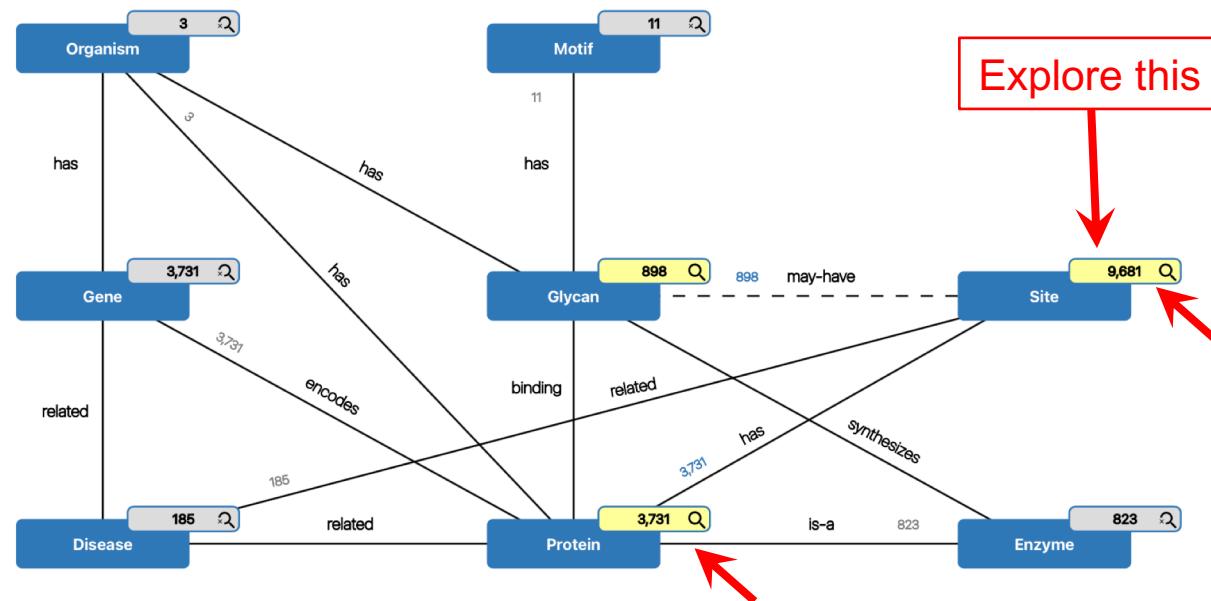
Show me all N-linked glycosylation sites that are impacted by a change in the Asn residue of the NXS/T/C sequon

Super Search

Super Search

Tutorial

Click on a node to get its properties.



Enable Debug

Simple View

Try Sample Query

Reset Query

Show Query

Site Search Summary

Performed on: 2021/04/29 22:51:05

Super search query.

[Update Results](#) [Modify Search](#)

** To perform the same search again using the current version of the database, click "Update Results".

Records per page 20 Showing 1 to 20 of 9681 Results

[DOWNLOAD](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [>](#) [>>](#)

UniProtKB Accession ↑↓	Hit Score ↑↓	Start Pos ↑↓	End Pos ↑↓	SNV	Glycosylation	Mutagenesis	Glycation	Phosphorylation
P27487-1	11.13	281	281	YES	YES	YES	NO	NO
Q9Y231-1	11.13	101	101	YES	YES	YES	NO	NO
Q15116-1	11.13	49	49	YES	YES	YES	NO	NO
Q10589-1	11.13	65	65	YES	YES	YES	NO	NO
Q02083-1	11.13	333	333	YES	YES	YES	NO	NO
Q8IWT6-1	11.13	66	66	YES	YES	YES	NO	NO
Q13510-1	11.13	348	348	YES	YES	YES	NO	NO
O60635-1	11.13	178	178	YES	YES	YES	NO	NO
Q9NUN5-1	11.13	448	448	YES	YES	YES	NO	NO
Q9H306-1	11.13	110	110	YES	YES	YES	NO	NO
Q9BXJ7-1	11.13	35	35	YES	YES	YES	NO	NO
Q8NCC3-1	11.13	398	398	YES	YES	YES	NO	NO
Q5VUB5-1	11.13	190	190	YES	YES	YES	NO	NO
Q8IWT6-1	11.13	83	83	YES	YES	YES	NO	NO

Explore this

Feedback

General
Glycosylation
Names
Function
Sequence
Single Nucleotide Variation
Mutagenesis
GO Annotation
Glycan Ligands
PTM Annotation
Proteoform Annotation
Pathway
Synthesized Glycans
Isoforms
Homologs
Disease
Expression Tissue
Expression Disease
Cross References
History
Publications

LOOK AT
Details For Glycoprotein **P27487-1**

Back

 DOWNLOAD ▾

Feedback

Explore this

General

Gene Name: DPP4

Gene Location: Chromosome: 2 (162,074,542 - 161,992,241)

Ensembl Gene 

UniProtKB ID: DPP4_HUMAN

UniProtKB Accession: P27487-1

Protein Length: 766

UniProtKB Entry Name: Dipeptidyl peptidase 4 membrane formDipeptidyl peptidase 4 soluble formDipeptidyl peptidase 4

Chemical Mass: 88,279 Da

RefSeq Accession: NP_001926.2

RefSeq Name: dipeptidyl peptidase 4 isoform 1

Organism: Homo sapiens (Human) [9606]

UniProtKB 

Glycosylation

 ProtVista

General
Glycosylation
Names
Function
Sequence
Single Nucleotide Variation
Mutagenesis
GO Annotation
Glycan Ligands
PTM Annotation
Proteoform Annotation
Pathway
Synthesized Glycans
Isoforms
Homologs
Disease
Expression Tissue
Expression Disease
Cross References
History
Publications

Sequence

ProtVista

+10 +20 +30 +40 +50

1 MKTPWKVLLG LLGAAALVTI ITPVVLNNK GT¹DATADSR KTYLTDYLK NTYRLKLYSL
 61 RWISDHEYLK KQ²NNILWVN AEYGN³SVFL ENST⁴DEFGH SINDYSIPD GQFILLEVN
 121 VQKWRHSYTA SYDIYDLNKK QLITEERIPM NT⁵WTWSPV GHKLAVWNN D⁶YVLEPNL
 181 PSYRITWTGK EDIYNGITD WYYEEEVFA YSALWSPNG TFLAYA⁷QND TEVPL⁸EYST
 241 YSD⁹ESLOQPK TVRVPPKAG AVNTVKFFV WNTDSLSSVT NATS¹⁰QITAP ASMLIGDHYL
 301 CD¹¹VWATQER ISLQLWLRRIQ NY¹²SMDICDY DESSGR¹³NCL VARQHIM¹⁴T TGWVGRFRPS
 361 EPHFTLDGNS FYKIISNEEG YRHICYFQID KKDC¹⁵FTIKG TWEVIGIEAL TSDYLYYISN
 421 EYKGMPGGRN LYKIQ¹⁶LSDYT KVTC¹⁷LSCELN PERCQYYSVS FSKE¹⁸AKYQL RCSGPGLPLY
 481 TLHSSVNDKG LRVLEDDNSA¹⁹ DKMLQNVQMP SKKLD²⁰FILN ETKF²¹WYQML PPHFDKSKKY
 541 PLLLDVYAGP CSQKADTVFR LNWATYLAST ENIVIASF²²D RGSGYQGD²³K MHAINRRLGT
 601 FEVEDQIEAA RQFSKMGFVD NKRIA²⁴IWGWS YYGYVTS²⁵MVL GSGSGVFKCG IAV²⁶PVSN²⁷WE
 661 YYDSVYTERY MGLPTPEDNL DHYRN²⁸WMS RAENFKOVEY LLIHG²⁹TADDN VHFOQS³⁰AQIS
 721 KALVDVGVD³¹ QAMWYTDEH GIASSTAHQH IYTH³²MSHFIK QCFSLP

N-linked Sites 12

O-linked Sites

Variation from mutation 32

Sequon 9

Explore this

Single Nucleotide Variation

ProtVista

Disease associated Mutations Non-disease associated Mutations

Records per page 20 Showing 1 to 19 of 19 Results

1

Source	Chromosome	Position	Ref	Alt	Type	Diseases
dbSNP 1	Chr2:162894910	172	172	I → S	kidney cancer (DOID:263)	
ExAC 1	Chr2:162891766	228	228	F → L	breast cancer (DOID:1612)	
TOPMed 1	Chr2:162891760	230	230	D → N	stomach cancer (DOID:10534) uterine cancer (DOID:362)	
gnomAD 1						
BioMut _n 1	Chr2:162894910	172	172	I → S	kidney cancer (DOID:263)	
BioMut _n 1	Chr2:162891766	228	228	F → L	breast cancer (DOID:1612)	
BioMut _n 1	Chr2:162891760	230	230	D → N	stomach cancer (DOID:10534) uterine cancer (DOID:362)	

Feedback

Invitation to Explore Features, Provide Feedback, Contact with Questions

GlyGen

HOME EXPLORE ▾ QUICK SEARCH TRY ME DATA ▾ TOOLS ▾ HELP ▾ MORE ▾

Search...



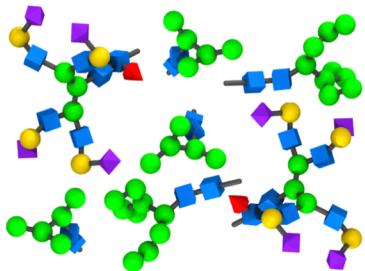
GlyGen: Computational and Informatics Resources for Glycoscience

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

HOW TO CITE

QUICK START

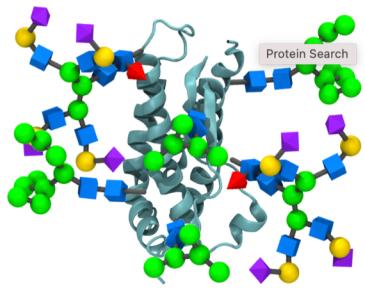
OUR MISSION



Glycan Search

Search for glycan structures based on their chemical and structural properties.

[EXPLORE](#)



Protein Search

Search for proteins based on their sequences, accessions, and annotations.

[EXPLORE](#)



Site Search

Search for proteins based on their site and site annotations.

[EXPLORE](#)

Version

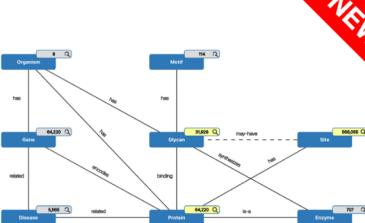
Portal: 1.8 (19/Apr/2021)
Webservice: 1.8.48 (22/Apr/2021)
Data: 1.8.25 (19/Apr/2021)



Quick Search

Quick Search provides multi-domain queries that are based on user requests.

[EXPLORE](#)



Super Search

Super search is a graphical interface to build queries across all GlyGen datasets.

[EXPLORE](#)

GlyGen Mapper

ID mapping related to glycan, protein / glycoprotein and based on the user input.

[EXPLORE](#)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

Database Statistics

Homo sapiens

Glycans	15855
Proteins	20609
Glycoproteins	9199

Mus musculus

Glycans	5842
Proteins	21989
Glycoproteins	3891

Rattus norvegicus

Glycans	4091
Proteins	21587
Glycoproteins	2111

Severe acute respiratory syndrome coronavirus 2

Glycans	1569
Proteins	17

Future plans, data integration, feature development:

- **New data types:** additional species, natural variants, phenotypes, phosphorylation, glycan binding protein interactions, protein-protein interactions, protein-GAG interactions, more glycosylation sites through automated literature mining
- **New annotations:** PTM functional annotations, more glycosylation subtypes, glycan names and relations (subsumption), biosynthetic pathways mapping enzymes to glycan precursor and products
- **New features:** enhanced filtering and sorting options, highlight hot-topics in glycoscience on the portal home page, new routes for accessing data through function/biosynthesis/disease, extensive help system using mediawiki, intuitive displays of data statistics, improved mobile friendliness